Responses to the Reviewer's Comments

Pan Jiang, Zhengjing Ma*, Gang Mei**
School of Engineering and Technology
China University of Geosciences (Beijing)
gang.mei@cugb.edu.cn

Acknowledgement The authors would like to thank the editor and the reviewer for their comments.

Comment # 1:

1. This is a well-structured and comprehensive review paper focusing on the application of deep learning (DL) in potential landslide identification. The manuscript systematically reviews data sources, model architectures, and representative applications, and it summarizes current challenges and future perspectives. The topic is timely and relevant to the research community given the rapid development of data-driven geohazard monitoring and early-warning systems. The revised version has clearly benefited from careful responses to the first-round comments-particularly through improved figure referencing, enriched citations, and enhanced case discussions in Sections 3 and 4.

Response:

- Dear reviewer, we sincerely thank you for the very positive and encouraging feedback on our revised manuscript! We are delighted that you find our review to be "well-structured and comprehensive, " and that the revisions made in response to the first-round comments have further strengthened the paper. Your feedback has been invaluable in strengthening the paper.

Comment # 2:

2. Please add a short quantitative statement in the abstract to highlight the scope of the reviewed literature. Example: "This review synthesizes more than 200 studies published between 2018 and 2025." This will increase the perceived depth of the review.

Response:

- Thank you for your insightful comment regarding the abstract! We appreciate your suggestion to add a quantitative statement to underscore the breadth of the literature covered.
- Accordingly, we have revised the abstract to explicitly state that this review synthesizes over 400 peer-reviewed studies, primarily published within the last six years, thereby highlighting both the scope and the timeliness of the review.

Original Description in Abstract

As global climate change and human activities escalate, the frequency and severity of landslide hazards have been increasing. Early identification, as an important prerequisite for monitoring, evaluation, and prevention, has become increasingly critical. Deep learning, as a powerful tool for data interpretation, has demonstrated remarkable potential in advancing landslide identification, particularly through the automated analysis of remote sensing, geological, and topographic data. This review provides an overview of recent advancements in the utilization of deep learning for potential landslide identification. First, the sources and characteristics of landslide-related data are summarized, including satellite observation data, airborne remote sensing data, and ground-based observation data. Next, commonly used deep learning models are classified based on their roles in potential landslide identification, covering areas such as image analysis and time series analysis. Then, the role of deep learning in identifying rainfall-induced landslides, earthquake-induced landslides, human activity-induced landslides, and multi-factor-induced landslides is summarized. Although deep learning has achieved considerable success in landslide identification, it still faces several challenges, including data imbalance, limited model generalization, and the inherent complexity of landslide mechanisms. Finally, future research directions in this field are discussed. It is suggested that integrating knowledge-driven and data-driven approaches for potential landslide identification will further enhance the applicability of deep learning, offering broad prospects for future research and practice.

Revised Description in Abstract

As global climate change and human activities escalate, the frequency and severity of landslide hazards have been increasing. Early identification, as an important prerequisite for monitoring, evaluation, and prevention, has become increasingly critical. Deep learning, as a powerful tool for data interpretation, has demonstrated remarkable potential in advancing landslide identification, particularly through the automated analysis of remote sensing, geological, and topographic data. This review systematically examines and synthesizes over 400 studies, with a primary focus on literature from the last six years (2020-2025), alongside key foundational works. It provides a comprehensive overview of recent advancements in the utilization of deep learning for potential landslide identification. First, the sources and characteristics of landslide-related data are summarized, including satellite observation data, airborne remote sensing data, and ground-based observation data. Next, commonly used deep learning models are classified based on their roles in potential landslide identification, covering areas such as image analysis and time series analysis. Then, the role of deep learning in identifying rainfall-induced landslides, earthquake-induced landslides, human activity-induced landslides, and multi-factor-induced landslides is summarized. Although deep learning has achieved considerable success in landslide identification, it still faces several challenges, including data imbalance, limited model generalization, and the inherent complexity of landslide mechanisms. Finally, future research directions in this field are discussed. It is suggested that integrating knowledge-driven and data-driven approaches for potential landslide identification will further enhance the applicability of deep learning, offering broad prospects for future research and practice.

Comment # 3:

3. Provide a comparative summary table linking DL architectures (CNN, RNN, Transformer, GAN, etc.) with specific landslide types and datasets (InSAR, optical, UAV, LiDAR). This will better highlight the methodological correspondence and identify research gaps.

Response:

- We sincerely thank you for this excellent and constructive suggestion! We fully agree that a comparative summary table will significantly enhance the synthesis and analytical depth of our review by clearly mapping the relationships between data source, deep learning models, and applications.
- Following your suggestion, we have added a new comparative summary table titled **"Typical correspondences among data source, deep learning models, and applications in potential landslide identification"** (now Table 1). To ensure it is placed where it offers the greatest integrative value, we have inserted this table at the end of Section 4 (Deep Learning for Potential Landslide Identification: Applications), specifically between Subsections 4.4 and 4.5. This location was chosen because Section 4 synthesizes the application outcomes of the models introduced in Section 3 and the data source detailed in Section 2. Placing the table here allows it to serve as a concluding synthesis of the entire "data-models-applications" pipeline, providing readers with a clear reference framework just before the section summary. Please see the revised **Section 4** for all details.
- This table consolidates and synthesizes the relationships previously discussed across Sections 2, 3, and 4, providing a concise overview of typical input data, target landslide types, and representative research tasks for each deep learning model. The detailed comparisons and mappings can be found in the **New Table 1**. We believe that this addition significantly enhances the clarity and integrative value of the review.

New Table 1

Deep Learning Models	Typical Input Data	Target Landslide Types	Representative Research Tasks
CNNs	Optical remote sensing imagery, UAV imagery, LiDAR-derived DEMs, and InSAR-derived deformation maps	Shallow landslides, rockfalls, and debris flows (with emphasis on morphological identification)	Landslide boundary delineation, susceptibility mapping, landslide inventory compilation, and pixel-level semantic segmentation
RNNs	InSAR time-series data and ground-based monitoring data (e.g., rainfall sequences, groundwater levels)	Creeping landslides and slow-moving landslides (focusing on time-series analysis)	Displacement prediction, temporal deformation analysis, and early warning systems
Transformers	Multi-temporal optical imagery, multi-sequence InSAR data, and multi-source environmental factors	Complex and multi-type landslides (particularly suitable for multi-source data fusion)	Multimodal landslide detection, change detection, and cross-domain prediction
GANs	Optical and UAV imagery, LiDAR-derived DEMs, and synthetic or augmented remote sensing data	Applicable across different landslide types (primarily used for data generation)	Data augmentation, sample generation, image reconstruction, and resolution enhancement
AEs	InSAR-derived surface deformation time series and high-dimensional multi-source landslide-related variables	Applicable across different landslide types (primarily used for feature learning and dimensionality reduction)	Feature extraction, anomaly detection, noise suppression, and dimensionality reduction
GNNs	Graph-structured spatial data derived from terrain units, sensor networks, or landslide inventories	Regional landslide systems, clustered landslides, and interacting slope units	Spatial interaction modeling, landslide clustering analysis, and network-based susceptibility analysis
Diffusion Models	Multi-source remote sensing data and synthetic datasets	Currently dominated by exploratory and methodological studies	Data denoising, generative modeling, uncertainty representation, and reconstruction

Original Description in Section 4

……

4.4 Application of Deep Learning in the Identification of Multi-factor-induced Landslides

……

4.5 Summary on the Applications of Deep Learning for Potential Landslide Identification

……

Revised Description in Section 4

……

4.4 Application of Deep Learning in the Identification of Multi-factor-induced Landslides

……

The diverse applications discussed in this section demonstrate that the selection and effectiveness of a deep learning model are fundamentally governed by the interplay between available data types, inherent model capabilities, and specific task objectives. To synthesize these critical relationships and provide a clear reference framework, Table 1 maps the typical correspondences between predominant deep learning architectures, their compatible data source, suited landslide phenomena, and representative application tasks. This synthesis underscores that there is no universally optimal model; rather, a strategic alignment across the data-model-application pipeline is key to successful implementation.

4.5 Summary on the Applications of Deep Learning for Potential Landslide Identification

……

Comment #4:

4. The restructured Section 3 is much clearer than before, yet it still reads like a tutorial. Consider emphasizing how these DL models have improved landslide detection or forecasting relative to traditional machine-learning or physically-based models.

Response:

- Thank you for accurately identifying the core weakness of our manuscript and for providing a clear direction for improvement! We agree that emphasizing the concrete advancements brought by deep learning is essential. In direct response to this comment, we have undertaken a substantive revision of Section 3 to explicitly shift its focus away from a tutorial-style presentation.
- The key revisions are summarized as follows:
- (1) We have significantly deleted and condensed the extensive descriptions of model architectures and operational details that contributed to the tutorial tone.
- (2) We have completely restructured the narrative logic for each deep learning category. The

discussion now consistently begins by identifying specific limitations in conventional machine learning or physics-based methods for a given task (e.g., manual feature engineering, difficulty in processing raw imagery or sequential data).

- (3) Building on that context, for each architecture, we now explicitly and foreground how it addresses those limitations and quantifiably improves performance metrics (e.g., detection accuracy, segmentation precision, predictive capability) as demonstrated in key comparative literature.
- We believe these revisions shift **Section 3** from a methodological tutorial toward a performance-oriented synthesis, thereby directly addressing your concern.

Original Description in Section 3

3.1 Models for Image Analysis and Processing in Potential Landslide Identification

Image data plays a critical role in potential landslide identification, especially through remote sensing, satellite, and UAV imagery. These images enable the acquisition of large-scale terrain data, encompassing complex geographical features, vegetation coverage, and ground fissures, which often serve as potential precursors to landslide occurrences. The adoption of deep learning has facilitated a shift from conventional manual visual interpretation to automated high-precision segmentation.

As illustrated in Fig. 2, a CNN is mainly composed of convolutional, pooling, and fully connected layers, each responsible for distinct operations on the input data (Kattenborn et al., 2021; LeCun et al., 1998; Liu et al., 2022b).

Convolutional layers, the core of CNNs, use kernels of various sizes to extract multi-scale features from geospatial imagery, which is crucial for landslide identification (Hussain et al. 2019; Shi et al. 2020; Yao et al. 2021). Small kernels are effective in detecting fine-grained precursors such as ground fissures and localized soil texture changes. For instance, Hamaguchi et al. (2018) proposed a Local Feature Extraction (LFE) module to enhance the capability of CNNs in identifying small object instances in remote sensing imagery. Wang et al. (2024a) demonstrated the exceptional capability of convolutional layers in extracting extremely small and subtle features by identifying cracks as narrow as 0.05 m width using a U-Net–based model. In contrast, larger kernels help in recognizing the overall morphology and boundaries of landslide bodies. From the perspective of general visual tasks, Ding et al. (2022) demonstrated that larger convolution kernels substantially improve the shape bias of CNNs, facilitating the recognition of large-scale structures and overall morphological patterns compared with using small kernels alone. Li et al. (2025) employed multiple large convolution kernels (kernel sizes = 5, 7, and 9) within the deep learning-based feature fusion with scale-adaptive kernel attention module to fuse multi-scale features, thereby enhancing the global perception of landslide boundaries and morphology as well as the capture of contextual background information.

Pooling layers down-sample feature maps, improving computational efficiency and model robustness. In landslide mapping, this translation invariance is particularly beneficial, as it allows the model to consistently identify landslide features regardless of their slight positional variations across different image patches (Mao et al., 2024).

The final fully connected layer flattens the pooled feature maps and performs classification, outputting results that distinguish potential landslide areas from non-landslide areas or enable further analysis of landslide types (Wu et al., 2024b).

The layers of a CNN can be combined in various ways, forming distinct CNN architectures. These architectures are primarily determined by task requirements, which may include image classification, multi-class segmentation, or object localization within a scene.

Conventional CNNs typically consist of multiple stacked convolutional layers, pooling layers, and fully connected layers. However, increasing network depth introduces challenges such as vanishing gradients and degradation arise, resulting in model performance deterioration.

ResNet mitigates the vanishing gradient problem in very deep networks through residual connections (Qi et al., 2020; Yang et al., 2022). This architectural advancement has been successfully applied to landslide detection in complex terrains, such as the work by Ullo et al. (2021), who demonstrated that a ResNet-based classifier could achieve high accuracy in distinguishing landslide scars from surrounding vegetation and bare soil in satellite imagery by effectively learning hierarchical features.

Models with higher parameter counts generally exhibit greater representational capacity but are prone to overfitting, while demanding higher computational resources and temporal costs for both training and inference (Ebrahimi and Abadi, 2021). For instance, He et al. (2016) introduced ResNet-152 and other deep residual network architectures, demonstrating that deeper structure achieve superior performance compared with shallower counterparts. Hasanah et al. (2023) explicitly highlighted the differences in layer depth and parameter count among various ResNet versions (ResNet-50, 101, and 152), noting that the increased number of parameters in deeper networks inevitably leads to longer training times.

DenseNet is a further innovation of ResNet (Huang et al., 2017). Both of these neural networks are based on a similar idea, which is to establish a "shortcut" between different layers. However, the structure of DenseNet is simpler and more effective, with fewer parameters. The structural differences between ResNet and DenseNet are illustrated in Fig. 2. In ResNet, each layer is only connected to the previous layer, while in DenseNet, each layer is directly connected to all previous layers, and each layer can obtain gradients from the loss function. This can optimize the information flow and gradients of the entire network, making it easier to train and performing better on small datasets. The structure of DenseNet enables more effective reuse of features, meaning that each layer can directly access and build upon the feature maps generated by all preceding layers instead of relearning similar representations. This dense connectivity not only strengthens information and gradient flow across the network but also reduces redundancy and the total number of parameters. Moreover, the layers of DenseNet are narrower than those of other deep learning networks (Liu et al., 2021c), making it reduce redundancy by learning with fewer feature maps. This architecture is suitable for the extraction of multi-scale landslide features under complex terrains, even with limited landslide training samples (Cai et al., 2021; Li et al., 2021; Ullo et al., 2021).

With the rapid expansion of deep learning methods based on CNNs, semantic segmentation models have increasingly become the standard in landslide detection (Lu et al., 2023b; Zhou et al., 2024b). As a fundamental task in computer vision, semantic segmentation assigns a specific class label (e.g., "landslide" or "non-landslide") to each pixel in an image, thereby enabling dense pixel-level classification (Guo et al., 2018).

Numerous advanced semantic segmentation networks have been proposed and validated for automatic landslide detection, significantly enhancing the efficiency and accuracy of large-scale detection.

U-Net is a typical example, which features a U-shaped architecture (Ronneberger et al., 2015). U-Net's encoder-decoder structure with skip connections has become a benchmark for landslide segmentation (Chandra et al., 2023; Chen et al., 2022b; Meena et al., 2022). For example, Nava et al. (2022) applied the attention U-Net to Sentinel-1 SAR data for rapid mapping of earthquake-induced landslides, demonstrating the effectiveness of U-Net variants in pixel-level segmentation of landslide bodies under cloud-covered or topographically complex conditions.

When dealing with complex features in landslide-prone areas, DeepLab is a more suitable choice than U-Net (Sandric et al., 2024). While U-Net excels at preserving fine-grained spatial details through its skip-connections, its ability to capture long-range contextual information is limited by its relatively small receptive field. DeepLab, built upon deep CNNs, addresses this critical limitation by employing dilated convolutions to exponentially expand the receptive field without sacrificing resolution or increasing parameters substantially.

More importantly, DeepLab integrates an Atrous Spatial Pyramid Pooling (ASPP) module, which is key to its superior performance on multi-scale objects like landslides (Chen et al., 2017; Huang et al., 2024a). The ASPP module operates in parallel on the same feature map using multiple convolutional branches with different dilation rates (e.g., rates = 6, 12, 18). Each branch effectively captures contextual information at a different scale, from fine details to broad, image-level contexts (Niu et al., 2018). All these multi-scale features are then concatenated and fused. This allows the network to simultaneously leverage both local textual cues and global contextual cues, thereby significantly improving recognition accuracy and reducing false positives in geologically complex environments.

After achieving semantic segmentation to obtain the accurate extent of a landslide and the classification of ground objects, change detection is employed to monitor the changes in the landslide area over time. By comparing the segmentation results of multiple temporal phases or directly analyzing the feature differences, the dynamic evolution of potential hazards can be quantified (Amankwah et al., 2022).

Wang (2023) demonstrates that 3D CNNs can directly process these 3D tensors. These models capture both spatial and temporal dependencies through 3D convolutional kernels, enabling the direct processing of multi-temporal image sequences. The outputs typically take two complementary forms: (1) change hotspot maps, which highlight regions of significant spatial change across time, and (2) temporal variation curves, which illustrate the evolution of pixel- or region-based feature values throughout the temporal sequence. Together, these representations provide intuitive and complementary tools for characterizing dynamic processes in landslide-prone areas, such as the initiation, progression, and spatial distribution of slope failures.

Some studies even have integrated attention mechanisms into conventional CNN architectures to enhance the analysis of multi-temporal remote sensing imagery, thereby enabling the identification of landslide hazard evolution over time. For example, Meng et al. (2024) proposed a framework based on CNN and optimized Bidirectional Gated Recurrent Unit (BiGRU) with an attention mechanism, designed to forecast landslide displacement with a step-like curve. Dong et al. (2022) proposed L-Unet which combines multi-scale feature fusion with attention modules to improve landslide segmentation performance, particularly at boundaries.

3.2 Models for Time Series Analysis in Potential Landslide Identification

The occurrence of a landslide is a gradual accumulation process, usually influenced by a variety of factors. We refer to data that reflect the changing states of a landslide body over time as time series data. Time series data analysis aims to excavate the information hidden in the time series data to help identify potential landslides.

Different from conventional statistical or physical models, deep learning models can automatically reveal dynamic change trends and periodic patterns in the data, providing more accurate information for landslide prediction and early warning. Recently, deep learning–based temporal models have become key tools for extracting nonlinear dependencies and temporal evolution patterns in landslide-related time series. The structural characteristics and differences among these models are illustrated in Fig. 3.

RNNs are a class of deep learning models specialized in processing sequential data, capable of capturing temporal dependencies within input sequences (Elman, 1990). Unlike conventional feedforward neural networks, in an RNN, each neuron not only receives the current input but also the output of the previous time step as additional input. This structure endows the RNN with a memory mechanism (Ngo et al., 2021; Zaremba et al., 2014).

In landslide prediction, RNNs have been employed to model displacement time series under rainfall or groundwater fluctuations, revealing short-term deformation patterns preceding slope failure (Chen et al., 2015; Zhang et al., 2022c).

To overcome the vanishing gradient problem inherent in RNNs, LSTM introduces memory cells and gating mechanisms that selectively retain relevant temporal information (Graves, 2012; Landi et al., 2021; Sherstinsky, 2020; Smagulova and James, 2019; Staudemeyer and Morris, 2019; Yu et al., 2019). As shown in Fig. 3, LSTM networks extend the basic RNN structure by incorporating gating units that control information flow, enabling them to better capture cumulative and delayed slope responses to environmental triggers. This capability allows them to model the cumulative and delayed responses of slopes to prolonged rainfall or reservoir water level fluctuations.

LSTM models have been widely applied in landslide displacement prediction and early warning. Yang et al. (2019) analyzed the relationships among landslide deformation, rainfall, and reservoir water levels, and found that compared with static models, the LSTM approach more accurately captured the dynamic characteristics of landslides and effectively leveraged historical information. Xu and Niu (2018) used a LSTM model to predict the displacement evolution of the Baijiabao landslide using rainfall and hydrological level data, achieving a higher correlation compared with traditional regression models. In another study focused on shallow landslides, Xiao et al. (2022) used a week-ahead LSTM model, which exhibited stable performance and improved prediction accuracy in short-term prediction scenarios. Additionally, Gidon et al. (2023) constructed a Bi-LSTM model and achieved a detection accuracy of 93% in the Mawiongrim area. This effectively addresses the limitations of traditional methods and can provide a reliable technical solution for disaster early warning in this area as well as other similar landslide-prone areas.

The GRU is a simplified variant of the LSTM that achieves similar accuracy with fewer parameters and reduced computational costs (Cho et al., 2014), making it well-suited for real-time landslide monitoring systems (Chung et al., 2014; Rawat and Barthwal, 2024; Zhang et al., 2022e).

Furthermore, GRU models effectively identify precursory displacement acceleration, allowing early detection of slope instability triggered by rainfall or seismic shaking (Chang et al., 2025; Yang et al., 2025).

Transformer, first introduced by Vaswani et al. (2017), was originally designed for natural language processing but has since become a cornerstone architecture in modern machine learning, achieving state-of-the-art performance across diverse domains such as computer vision and multimodal learning.

Unlike conventional recurrent or convolutional models, the Transformer is built upon stacked encoder–decoder layers and relies on a key innovation: the self-attention mechanism (see Fig. 5). This mechanism enables the model to automatically compute a weight vector (i.e., an attention distribution) for each element in the sequence based on its relevance to all other elements. By evaluating all positions simultaneously (Esser et al., 2021; Huang and Chen, 2023), the Transformer efficiently captures global dependencies across long sequences in parallel, making it more effective than RNNs or CNNs at modeling long-range relationships.

When applied to landslide-related time series data, the Transformer can adaptively learn latent temporal features and patterns, automatically adjusting parameters to accommodate diverse landslide scenarios (Wang et al., 2024b; Zerveas et al., 2021).

However, a key drawback of the standard Transformer is its quadratic computational complexity with respect to sequence length, which becomes prohibitive for very long sequences (Zhuang et al., 2023). This also complicates the interpretation of how the model extracts features and makes decisions from large amounts of landslide data, posing challenges for practical deployment. It is worth noting that mitigating this quadratic complexity is an active research area, with many efficient Transformer variants being developed. For example, Zhao et al. (2024f) combined the strengths of CNN and Transformer architectures, selecting and analyzing nine landslide conditioning factors to successfully achieve accurate landslide localization and detailed feature capture. Ge et al. (2024) proposed the LiteTransNet model based on the Transformer framework, effectively capturing and interpreting the varying importance of historical information during the prediction process. Therefore, while powerful, the vanilla Transformer may not be the optimal choice for all practitioners, and its computational demands should be carefully considered.

In contrast, RNN-based models exhibit a relatively simple architecture and are conceptually intuitive (Li et al., 2021; Wang et al., 2020b), making them more interpretable. Transformers, however, are structurally more complex with numerous parameters, requiring substantial computational resources during training and being susceptible to overfitting on small datasets.

3.3 Models for Data Generation in Potential Landslide Identification

Data generation refers to modeling the underlying data distribution to generate entirely new samples independent of the original dataset (Kingma et al., 2014; Moreno-Barea et al., 2020; Shorten and Khoshgoftaar, 2019), thereby enriching the dataset. In potential landslide identification, data generation mitigates challenges of data scarcity and imbalanced class distributions, thereby enhancing the generalization capability of predictive models.

Deep generative models are the leading deep learning approach for synthetic data generation (Alam et al., 2018; Karras et al., 2020; Ma et al., 2024; Xu et al., 2015). They utilize deep neural networks to learn latent representations of data and optimize the learning process through specific objective functions. A key characteristic of deep generative models lies in their probabilistic nature. They not only classify or reconstruct data but also capture the underlying distribution of geospatial features, thereby enabling the generation of new landslide samples that are statistically consistent with observed patterns. Commonly used deep generative models include GANs, Variational Autoencoders (VAEs), and diffusion models (see Fig. 4).

GANs consist of a generator and a discriminator that compete in an adversarial process (Goodfellow et al., 2014). The generator synthesizes data resembling real samples, while the discriminator attempts to distinguish between generated and real data. The workflow of adversarial training for GAN-based data generation is schematically depicted in Fig. 4. Through iterative adversarial training, the generator learns to produce high-quality synthetic data that closely matches the distribution of real data (Gui et al., 2021; Saxena and Cao, 2021).

In the context of landslide studies, GANs have demonstrated strong capabilities in data augmentation and remote sensing image enhancement. For example, Feng et al. (2024) achieved the first implementation of using a GAN to generate synthetic high-quality landslide images, aiming to address the data scarcity issue that undermines the performance of landslide segmentation models. Al-Najjar and Pradhan (2021b) proposed a novel approach that employs a GAN to generate synthetic inventory data. The results indicate that additional samples produced by the proposed GAN model can enhance the predictive performance of Decision Trees (DT), Random Forest (RF), Artificial Neural Network (ANN), and Bagging ensemble models.

Despite their advantages, GANs may suffer from mode collapse, leading to limited diversity in the generated data, especially when certain landslide types are underrepresented (Fang et al., 2020a). Moreover, their unstable training process requires careful hyperparameter tuning and substantial computational resources, which may constrain their application in real-time hazard scenarios. Nevertheless, with improved architectures such as Conditional GAN (CGAN) (Kim and Lee, 2020; Loey et al., 2020; Mirza and Osindero, 2014), image-to-image translation with GAN (Pix2Pix) (Isola et al., 2017; Qu et al., 2019), and Wasserstein GAN (WGAN) (Arjovsky et al., 2017; Wang et al., 2019), GANs are becoming increasingly viable tools for high-resolution landslide mapping and synthetic data generation in remote sensing-based susceptibility analysis.

As a probabilistic variant of AEs, VAEs introduce latent-space regularization through variational inference (Hinton and Salakhutdinov, 2006; Kingma andWelling, 2013). The encoder compresses input data into a latent representation characterized by a mean and a standard deviation, while the decoder reconstructs the data by sampling from this distribution. This enables the model to generate new data with inherent randomness and diversity (Islam et al., 2021; Oliveira et al., 2022).

In landslide research, VAEs have been successfully applied to learn and reconstruct geomorphological patterns of slope instability. For instance, Cai et al. (2024) proposed and demonstrated the superior capability of the VAE-GRU model in generating narrow predictive intervals while maintaining high coverage probabilities, representing a substantial improvement over the state-of-the-art methods for probabilistic landslide prediction.

Compared with GANs, VAEs produce more diverse but slightly less detailed samples, due to their structured latent space constraints. This characteristic is particularly beneficial for exploring a wide range of potential landslide morphologies and for augmenting training datasets used in susceptibility prediction. However, VAEs may still struggle with highly imbalanced datasets, as their probabilistic reconstruction tends to favor majority classes. Integrating VAEs with stratified sampling or cost-sensitive learning could help overcome this limitation and further enhance landslide prediction performance.

When computational resources and training time permit, diffusion models provide a powerful alternative for generating high-quality, diverse, and stable data (Croitoru et al., 2023; Ho et al., 2020; Yang et al., 2023a; Zhu et al., 2023a). These models learn the data distribution by gradually adding noise to real samples (forward diffusion) and then reconstructing clean data through a reverse denoising process (see Fig. 4). The resulting models can sample new, realistic data points that reflect complex terrain and geophysical variability. For example, Lo and Peters (2024) proposed a Terrain-Feature-Guided Diffusion Model (TFDM) to fill gaps in DEM data. Similarly, Zhao et al. (2024b) employed a Denoising Diffusion Probabilistic Model (DDPM) conditioned on incomplete DEMs, which serves as a transitional kernel during diffusion reversal to progressively reconstruct sharp and accurate DEMs.

Despite their successful applications in image synthesis, denoising, and remote-sensing image enhancement (Leher et al., 2025; Sui et al., 2024; Xiao et al., 2023; Zou et al., 2024), diffusion models have not yet been widely applied directly to the identification of potential landslides and remain in the exploratory stage. Nonetheless, our optimism for their application is grounded in their potential to address key challenges such as limited labeled data through generative augmentation and, more importantly, to provide uncertainty quantification in predictions, which is vital for risk assessment.

In conclusion, deep generative models provide a transformative solution for overcoming the challenges of limited and imbalanced landslide datasets. By synthesizing realistic, diverse, and statistically consistent samples, these models can improve the robustness and generalization of landslide prediction frameworks. Among them, GANs are effective for generating visually realistic imagery and data augmentation; VAEs capture probabilistic geomorphic transitions; and diffusion models ensure stability and fidelity in high-resolution terrain synthesis.

3.4 Models for Anomaly Detection in Potential Landslide Identification

Anomaly detection plays a critical role in potential landslide identification, as it enables the distinction between normal environmental variations and genuine precursors of slope instability (Deijns et al., 2020; Jiang et al., 2020). In landslide monitoring, the goal of anomaly detection is to identify subtle yet significant deviations. Examples include abnormal surface displacements, changes in surface coherence, or irregularities in sensor signals. Such deviations may occur prior to landslide events. With the advancement of deep learning, data filtering has evolved from rule-based threshold detection to automated feature learning, allowing models to capture complex spatiotemporal dependencies and identify anomalies within high-dimensional, multi-source datasets.

AEs are widely used for unsupervised anomaly detection due to their ability to reconstruct input data and highlight deviations from learned normal patterns (Sakurada and Yairi, 2014; Zhou and Paffenroth, 2017). An AE consists of an encoder that compresses data into a low-dimensional latent representation and a decoder that reconstructs it.

During training, the AE learns the intrinsic features of normal landslide data, such as sensor-based displacement time series or radar backscatter from stable slopes. When abnormal data are input, such as sudden displacement spikes or incoherent radar signals, the reconstruction error increases significantly, serving as an indicator of potential instability. For instance, Shakeel et al. (2022) developed an InSAR deformation anomaly detector based on an AE–LSTM architecture. Experimental analyses using synthetic deformation test scenarios achieved an overall performance accuracy of 91.25%.

By defining a reconstruction error threshold, anomalies can be quantitatively detected. When the reconstruction error of new sensor data exceeds this threshold, it may signal slope movement acceleration or surface disturbance associated with potential landslides. Thus, AEs provide a data-driven method to detect early-warning signs without requiring manually labeled failure data.

As previously introduced, VAE is a probabilistic extension of AEs (Nawaz et al., 2024). VAEs introduce stochastic latent variables characterized by mean and variance, allowing them to model data uncertainty (see Fig. 4). During training, VAEs learn the latent distribution of normal samples and reconstruct inputs accordingly. When new observation data deviate significantly from the learned distribution, the reconstruction error increases accordingly, and this phenomenon can be used as an indicator of potential anomalies (Kingma and Welling, 2013; Li et al., 2020; Park et al., 2018).

In landslide applications, VAEs have been shown to outperform conventional AEs in handling complex, multivariate datasets that integrate topographic, meteorological, and geotechnical factors. For example, Han et al. (2025) proposed an unsupervised failure mode recognition algorithm based on a deep convolutional autoencoder, which integrates surface displacement, vertical displacement, and rainfall monitoring data from slopes to accurately identify the developmental stages of slope failure, achieving a recognition accuracy of 99.30%.

Compared to AEs, VAEs are particularly advantageous for capturing uncertainty and latent correlations between environmental variables, making them ideal for anomaly detection in integrated landslide early-warning systems (Kumar et al., 2024; Pol et al., 2019). However, they require larger datasets for stable training, and their probabilistic outputs may demand postprocessing for operational thresholding.

GANs can also be adapted for anomaly detection by exploiting their discriminator network's ability to distinguish between real and generated data (Kang et al., 2024; Xia et al., 2022). In landslide monitoring, GAN-based anomaly detection models learn the distribution of stable slope features, and deviations from this distribution can indicate abnormal conditions (Radoi, 2022).

AnoGAN extends conventional GANs by directly incorporating anomaly detection as one of its primary objectives (Lin et al., 2023; Thomine et al., 2023). It introduces an additional encoder during training, which maps input data to the latent space. The difference between this latent vector and the latent vector of normal samples generated by the generator serves as the basis for anomaly detection.

RNNs and their variants are particularly effective for time series–based anomaly detection, learning temporal dependencies and predicting future trends (Zamanzadeh Darban et al., 2024; Zhang et al., 2022a). In landslide monitoring, these models can process continuous displacement or rainfall time series to identify deviations from expected temporal behavior. These temporal models complement image-based approaches by providing continuous surveillance and early detection capabilities (Wu et al., 2024a).

When combined with AEs or GANs, RNN-type architectures can form hybrid frameworks capable of both spatial and temporal anomaly detection, enabling multi-source consistency checking in landslide early-warning systems. Geiger et al. (2020) demonstrated a growing trend of utilizing LSTM networks as both the generator and discriminator within GAN frameworks for time-series anomaly detection. Similarly, Whitaker (2023) illustrated the application of LSTM–GAN architectures in identifying temporal anomalies.

3.5 Models for Data Fusion in Potential Landslide Identification

In practical applications, the identification of potential landslide hazards is a complex task that influences by multiple factors (Zhang et al., 2018). These factors are often reflected through different data sources. We can roughly divide heterogeneous data into four categories: image data, time series data, structured data, and textual data. Given this heterogeneity, data fusion is essential for the accurate identification of potential landslides.

Since heterogeneous data differ in feature scale, spatial resolution, and data modality, deep learning models are increasingly utilized to automatically extract nonlinear and high-order feature interactions across data sources, offering significant advantages over conventional statistical fusion techniques. In landslide applications, deep learning-based data fusion can integrate multi-modal inputs such as Sentinel-1 InSAR deformation, rainfall time series, and terrain derivatives for regional-scale susceptibility mapping or real-time early warning.

Due to the non-Euclidean and topologically complex nature of landslide-related terrain, conventional CNN-based models are limited in representing irregular spatial dependencies. Graph Neural Networks (GNNs) have emerged as powerful architectures to model such relationships by representing spatial entities (e.g., slope units, grid cells, or sensor nodes) as graph nodes and their geospatial or topological interactions as edges (Scarselli et al., 2008; Ying et al., 2018; Zeng et al., 2022).

In landslide identification, GNNs enable explicit modeling of spatial connectivity and geological adjacency, allowing the propagation of geomorphic and hydrological information across neighboring units. For example, Kuang et al. (2022) proposed an innovative landslide forecasting model based on GNNs, in which graph convolutions are employed to aggregate spatial correlations among different monitoring sites. Ren et al. (2025) introduced a novel GNN framework with conformal prediction (GNN-CF) for landslide deformation interval forecasting, addressing the limitations of conventional models in handling predictive uncertainty.

According to the differences in message passing and aggregation methods, GNNs have derived various variants. For example, Graph Convolutional Network (GCN) is generated by generalizing the convolutional operation to graph-structured data (Kipf and Welling, 2016; Sharma et al., 2022; Wang et al., 2020a), and Graph Attention Network (GAT) dynamically weights the importance of neighboring nodes by introducing the attention mechanism (Veliˇckovi´c et al., 2017; Yuan et al., 2022; Zhou and Li, 2021). The emergence of these new architectures makes GNN variants more targeted than conventional GNNs and suitable for modeling heterogeneous relationships. Currently, they are often used for weighted analysis of the impacts of different geographical factors on landslides (Kuang et al., 2022; Li et al., 2025; Zhang et al., 2024e).

As highlighted in Section 3.2, the Transformer's self-attention mechanism and modular architecture make it a universal framework for processing sequential data and enabling multimodal fusion (see Fig. 5).

In this context, the core advantage of the Transformer lies in its ability to integrate diverse input data (e.g., satellite imagery, GPS time series, and geological maps). It achieves this by employing independent embedding layers to convert each modality into a unified vector representation, which is then fused through the self-attention mechanism. This mechanism computes the interactions and correlations among all elements across different modalities, thereby enabling the model to capture cross-modal dependencies and extract joint feature representations within a unified framework. This capability makes the Transformer particularly suitable for landslide studies (Li et al., 2025). For example, Piran et al. (2024) enhanced short-term precipitation forecasting by applying transfer learning with a pre-trained Transformer model. Zhang et al. (2024e) incorporated Transformer modules to build a graph-Transformer model that integrates global contextual information for the generation and analysis of Landslide Susceptibility Maps (LSMs).

Revised Description in Section 3

3.1 Models for Image Analysis and Processing in Potential Landslide Identification

Image data plays a critical role in potential landslide identification, especially through remote sensing, satellite, and UAV imagery. These images enable the acquisition of large-scale terrain data, encompassing complex geographical features, vegetation coverage, and ground fissures, which often serve as potential precursors to landslide occurrences. The adoption of deep learning has facilitated a shift from conventional manual visual interpretation to automated high-precision segmentation.

CNNs, owing to their inherent capability to learn hierarchical and multi-scale spatial features (Kattenborn et al., 2021; LeCun et al., 1998; Liu et al., 2022b), have become the core methodological framework for most image-based deep learning applications in landslide research (see Fig. 2). This capability directly addresses a long-standing limitation of conventional classifiers, which struggle to simultaneously capture fine-scale precursors (e.g., narrow ground fissures) and large-scale landslide morphology within a unified framework. Multi-scale convolutional feature extraction has been shown to significantly enhance the sensitivity of landslide detection across a wide range of spatial extents (Hussain et al., 2019; Shi et al. 2020; Yao et al. 2021). For example, small convolutional kernels are particularly effective in identifying subtle surface disturbances, such as localized soil texture variations and ground cracks, which often precede slope failure. Hamaguchi et al. (2018) and Wang et al. (2024a) demonstrated that CNN-based models can detect extremely small and subtle features, including cracks as narrow as 0.05 m, a level of detail that is difficult to achieve using conventional texture-based methods.

Conversely, larger convolutional kernels and multi-scale fusion strategies enhance the identification of overall landslide morphology and scar boundaries, which are critical for accurate inventory mapping. Ding et al. (2022) showed that larger kernels improve the shape bias of CNNs, facilitating the recognition of large-scale structural patterns, while Li et al. (2025) demonstrated that scale-adaptive kernel fusion improves global perception of landslide extents and contextual background information. By integrating multi-scale feature extraction within a single model, CNN-based approaches outperform conventional machine-learning classifiers that depend on fixed-scale descriptors and often exhibit reduced generalization in heterogeneous terrain.

Beyond feature extraction, architectural innovations such as residual and dense connections have substantially improved the trainability and data efficiency of deep networks in landslide applications (He et al., 2016). Deep networks with increased depth generally exhibit stronger representational capacity but are prone to optimization difficulties and overfitting, particularly under limited training samples (Ebrahimi and Abadi, 2021).

Residual Networks (ResNet) address these challenges through shortcut connections (Qi et al., 2020; Yang et al., 2022), enabling stable training of very deep models and improved discrimination between landslide scars and surrounding vegetation or bare soil in complex terrains (see Fig. 2c). However, deeper architectures also incur higher computational costs, which may constrain their practical deployment in large-scale or near-real-time mapping scenarios (Hasanah et al., 2023).

Dense Convolutional Networks (DenseNet) further enhance feature reuse and gradient flow through dense connectivity, reducing parameter redundancy and improving performance under limited training data conditions (Huang et al., 2017; Liu et al., 2021c). This property is particularly relevant for landslide studies, where high-quality labeled samples are often scarce and spatially clustered. Empirical studies indicate that DenseNet-based models can effectively extract multi-scale landslide features in complex terrain while maintaining computational efficiency (Cai et al., 2021; Li et al., 2021; Ullo et al., 2021).

With the maturation of CNN backbones, semantic segmentation has emerged as the dominant paradigm for landslide detection, as it enables dense, pixel-level delineation of landslide extents that is essential for inventory construction and hazard assessment (Guo et al., 2018; Lu et al., 2023b; Zhou et al., 2024b). Among these models, U-Net and its variants have become benchmarks due to their encoder–decoder structure and skip connections, which preserve spatial detail and improve boundary delineation (Chandra et al., 2023; Chen et al., 2022b; Meena et al., 2022; Ronneberger et al., 2015). U-Net-based models have demonstrated strong performance in challenging conditions, such as cloud-covered or topographically complex regions using SAR imagery (Nava et al. 2022).

However, U-Net's relatively limited receptive field can restrict its ability to capture long-range contextual information in heterogeneous geological settings. DeepLab addresses this limitation by incorporating dilated convolutions and Atrous Spatial Pyramid Pooling (ASPP), enabling effective fusion of local texture and global contextual cues without sacrificing spatial resolution (Chen et al., 2017; Huang et al., 2024a). This multi-scale contextual modeling has been shown to reduce false positives and improve detection consistency in geologically complex environments, highlighting a key advantage of advanced deep segmentation models over simpler pixel-based or object-based approaches (Niu et al., 2018; Sandric et al., 2024).

Beyond static mapping, deep learning also facilitates multi-temporal change detection and dynamic hazard monitoring. By comparing segmentation outputs across time or directly processing multi-temporal image stacks, CNN-based models can characterize the spatial evolution of landslides and identify active deformation zones (Amankwah et al., 2022). Wang (2023) demonstrates that 3D CNNs enable joint modeling of spatial and temporal dependencies, producing both change hotspot maps and temporal evolution curves that capture landslide initiation and progression. Some studies even have integrated attention mechanisms into conventional CNN architectures to enhance the analysis of multi-temporal remote sensing imagery, thereby enabling the identification of landslide hazard evolution over time. For example, Meng et al. (2024) proposed a framework based on CNN and optimized Bidirectional Gated Recurrent Unit (BiGRU) with an attention mechanism, designed to forecast landslide displacement with a step-like curve. Dong et al. (2022) proposed L-Unet which combines multi-scale feature fusion with attention modules to improve landslide segmentation performance, particularly at boundaries.

Overall, image-based deep learning models represent a substantial methodological advance over traditional machine-learning classifiers in terms of multi-scale feature representation, mapping completeness, and robustness to complex backgrounds. Nevertheless, their performance remains contingent on data quality, sample representativeness, and computational resources, and they generally lack the explicit physical interpretability of process-based models. These limitations motivate increasing interest in hybrid framework.

3.2 Models for Time Series Analysis in Potential Landslide Identification

Landslide occurrence is inherently a time-dependent process, driven by the cumulative and often delayed effects of environmental forcing such as rainfall, groundwater fluctuation, reservoir operation, and seismic disturbance. Time series data describing slope displacement, pore-water pressure, rainfall intensity, or surface deformation provide critical information for identifying potential instability and forecasting landslide evolution. Unlike static susceptibility mapping, time series analysis directly targets the dynamic behavior of slopes and therefore plays a central role in early warning and short-term prediction (see Fig. 3).

Conventional statistical and physically based approaches have been widely used to analyze landslide-related time series. Statistical models typically assume linear or weakly nonlinear relationships and often require strong prior assumptions, while physically based models rely on simplified representations of hydromechanical processes and detailed parameterization that is difficult to obtain at scale. Deep learning–based temporal models offer a complementary data-driven alternative by automatically learning nonlinear dependencies, cumulative effects, and delayed responses directly from observations, without requiring explicit process equations.

RNNs represent the earliest class of deep learning models designed for sequential data, enabling the modeling of short-term temporal dependencies through recursive information flow (Elman, 1990; Ngo et al., 2021; Zaremba et al., 2014). In landslide studies, RNNs have been applied to displacement time series influenced by rainfall and groundwater variation, demonstrating their ability to capture short-term deformation trends prior to failure (Chen et al., 2015; Zhang et al., 2022c). However, standard RNNs often struggle with long-term dependencies and cumulative effects, which are common in landslide processes driven by prolonged or intermittent forcing (see Fig. 3b).

To overcome the vanishing gradient problem inherent in RNNs, LSTM introduces memory cells and gating mechanisms that selectively retain relevant temporal information (Graves, 2012; Landi et al., 2021; Sherstinsky, 2020; Smagulova and James, 2019; Staudemeyer and Morris, 2019; Yu et al., 2019). This capability is particularly well aligned with landslide dynamics, where delayed and cumulative responses to rainfall or reservoir level fluctuations are critical precursors of instability. Empirical studies consistently demonstrate that LSTM-based models outperform conventional regression and shallow machine-learning approaches in displacement prediction and early warning tasks. For example, Yang et al. (2019) analyzed the relationships among landslide deformation, rainfall, and reservoir water levels, and found that compared with static models, the LSTM approach more accurately captured the dynamic characteristics of landslides and effectively leveraged historical information. Xu and Niu (2018) used a LSTM model to predict the displacement evolution of the Baijiabao landslide using rainfall and hydrological level data, achieving a higher correlation compared with traditional regression models. In another study focused on shallow landslides, Xiao et al. (2022) used a week-ahead LSTM model, which exhibited stable performance and improved prediction accuracy in short-term prediction scenarios. Additionally, Gidon et al. (2023) constructed a Bi-LSTM model and achieved a detection accuracy of 93% in the Mawiongrim area.

Despite their strong performance, LSTM models are computationally demanding and may be prone to overfitting when training data are limited. GRUs provide a streamlined alternative by simplifying the gating structure while maintaining comparable predictive accuracy (Cho et al., 2014). This balance between model complexity and performance makes GRU-based models particularly attractive for real-time landslide monitoring and operational early warning systems, where computational efficiency and rapid updating are critical (Chung et al., 2014; Rawat and Barthwal, 2024; Zhang et al., 2022e). Recent studies indicate that GRUs can effectively identify acceleration phases in displacement time series, enabling earlier detection of rainfall- or earthquake-triggered slope instability (Chang et al., 2025; Yang et al., 2025).

More recently, Transformer-based architectures have emerged as powerful alternatives for time series modeling by leveraging self-attention mechanisms to capture long-range temporal dependencies in parallel (Vaswani et al. 2017). Compared with recurrent models, Transformers are particularly effective at modeling long-term and non-local temporal relationships, which are often present in landslide processes influenced by multi-seasonal rainfall or complex hydrological regimes. In landslide-related applications, Transformers can adaptively learn latent temporal features across diverse scenarios and outperform conventional RNN-based models in capturing complex temporal patterns (Esser et al., 2021; Huang and Chen, 2023; Wang et al., 2024b; Zerveas et al., 2021).

However, a key drawback of the standard Transformer is its quadratic computational complexity with respect to sequence length, which becomes prohibitive for very long sequences (Zhuang et al., 2023). This also complicates the interpretation of how the model extracts features and makes decisions from large amounts of landslide data, posing challenges for practical deployment. It is worth noting that mitigating this quadratic complexity is an active research area, with many efficient Transformer variants being developed. For example, Zhao et al. (2024f) combined the strengths of CNN and Transformer architectures, selecting and analyzing nine landslide conditioning factors to successfully achieve accurate landslide localization and detailed feature capture. Ge et al. (2024) proposed the LiteTransNet model based on the Transformer framework, effectively capturing and interpreting the varying importance of historical information during the prediction process. Therefore, while powerful, the vanilla Transformer may not be the optimal choice for all practitioners, and its computational demands should be carefully considered.

In summary, deep learning-based time series models represent a significant advancement over conventional statistical approaches by enabling data-driven learning of nonlinear, delayed, and cumulative deformation patterns that are difficult to encode explicitly in physical models. RNNs and LSTMs remain effective and interpretable for short- to medium-term prediction tasks, while GRUs offer computationally efficient solutions for operational systems (Li et al., 2021; Wang et al., 2020b). Transformer-based models provide superior capacity for long-term dependency modeling but require careful consideration of data availability, computational resources, and interpretability. These trade-offs highlight the importance of selecting temporal architectures based on specific monitoring objectives, data characteristics, and operational constraints.

3.3 Models for Data Generation in Potential Landslide Identification

A fundamental challenge in potential landslide identification lies in the scarcity, imbalance, and spatial clustering of labeled landslide samples. Landslide inventories are often incomplete, biased toward large or easily detectable events, and unevenly distributed in space and time. These limitations significantly constrain the performance and generalization ability of both traditional machine-learning classifiers and deep learning-based models, particularly in data-hungry settings. Data generation aims to alleviate these issues by learning the underlying data distribution and synthesizing new samples that are statistically consistent with observed landslide patterns (Kingma et al., 2014; Moreno-Barea et al., 2020; Shorten and Khoshgoftaar, 2019).

Conventional data augmentation techniques (e.g., rotation, flipping, noise injection) provide limited diversity and do not fundamentally address class imbalance or morphological variability in landslide datasets. Deep generative models represent a major methodological advance by explicitly modeling the latent distribution of geospatial features, thereby enabling the creation of realistic and diverse synthetic landslide samples (Alam et al., 2018; Karras et al., 2020; Ma et al., 2024; Xu et al., 2015). Unlike discriminative models, generative models capture probabilistic representations of terrain, deformation, or image features, making them particularly suitable for addressing uncertainty, rarity, and heterogeneity in landslide data. Commonly used deep generative models include GANs, Variational Autoencoders (VAEs), and diffusion models (see Fig. 4).

GANs are among the most widely adopted generative models for landslide-related data augmentation, particularly in remote sensing imagery. Through adversarial training between a generator and a discriminator, GANs can produce visually realistic synthetic samples that closely resemble real landslide images (Goodfellow et al., 2014; Gui et al., 2021; Saxena and Cao, 2021). In potential landslide identification, this capability can address the shortage of labeled image samples that limits the performance of segmentation and classification models. For example, Feng et al. (2024) achieved the first implementation of using a GAN to generate synthetic high-quality landslide images, aiming to address the data scarcity issue that undermines the performance of landslide segmentation models. Al-Najjar and Pradhan (2021b) proposed a novel approach that employs a GAN to generate synthetic inventory data. The results indicate that additional samples produced by the proposed GAN model can enhance the predictive performance of Decision Trees (DT), Random Forest (RF), Artificial Neural Network (ANN), and Bagging ensemble models.

Despite their effectiveness, GAN-based approaches exhibit notable limitations. Mode collapse may reduce sample diversity, particularly for rare landslide types or extreme morphologies, and training instability often necessitates careful hyperparameter tuning and substantial computational resources (Fang et al., 2020a). Such constraints can limit their applicability in operational or real-time hazard assessment. Recent architectural refinements, including Conditional GAN (CGAN) (Kim and Lee, 2020; Loey et al., 2020; Mirza and Osindero, 2014), image-to-image translation with GAN (Pix2Pix) (Isola et al., 2017; Qu et al., 2019), and Wasserstein GAN (WGAN) (Arjovsky et al., 2017; Wang et al., 2019), partially mitigate these issues by improving training stability and enabling conditional or controlled sample generation. As a result, GANs are increasingly viable for high-resolution landslide image synthesis and remote sensing–based susceptibility analysis, particularly when visual realism is a primary requirement.

As a probabilistic variant of AEs, VAEs introduce latent-space regularization through variational inference (see Fig. 4c). Compared with GANs, VAEs prioritize distributional coverage and uncertainty representation over visual sharpness (Hinton and Salakhutdinov, 2006; Kingma andWelling, 2013), making them well suited for probabilistic modeling of landslide processes. For instance, Cai et al. (2024) demonstrated that a VAE–GRU framework can generate narrow predictive intervals while maintaining high coverage probabilities, representing a substantial improvement over the state-of-the-art methods. Such probabilistic outputs are particularly valuable for risk-informed decision-making and early warning applications (Islam et al., 2021; Oliveira et al., 2022).

Compared with GANs, VAEs produce more diverse but slightly less detailed samples, due to their structured latent space constraints. This characteristic is particularly beneficial for exploring a wide range of potential landslide morphologies and for augmenting training datasets used in susceptibility prediction. However, VAEs may still struggle with highly imbalanced datasets, as their probabilistic reconstruction tends to favor majority classes. Integrating VAEs with stratified sampling or cost-sensitive learning could help overcome this limitation and further enhance landslide prediction performance.

When computational resources and training time permit, diffusion models provide a powerful alternative for generating high-quality, diverse, and stable data (Croitoru et al., 2023; Ho et al., 2020; Yang et al., 2023a; Zhu et al., 2023a). The resulting models can sample new, realistic data points that reflect complex terrain and geophysical variability. For example, Lo and Peters (2024) proposed a Terrain-Feature-Guided Diffusion Model (TFDM) to fill gaps in DEM data. Similarly, Zhao et al. (2024b) employed a Denoising Diffusion Probabilistic Model (DDPM) conditioned on incomplete DEMs, which serves as a transitional kernel during diffusion reversal to progressively reconstruct sharp and accurate DEM.

Despite their successful applications in image synthesis, denoising, and remote-sensing image enhancement (Leher et al., 2025; Sui et al., 2024; Xiao et al., 2023; Zou et al., 2024), diffusion models have not yet been widely applied directly to the identification of potential landslides and remain in the exploratory stage. Nonetheless, our optimism for their application is grounded in their potential to address key challenges such as limited labeled data through generative augmentation and, more importantly, to provide uncertainty quantification in predictions, which is vital for risk assessment.

In summary, deep generative models provide an essential complement to discriminative deep learning and conventional machine-learning approaches in potential landslide identification. Among them, GANs are effective for generating visually realistic imagery and data augmentation; VAEs capture probabilistic geomorphic transitions; and diffusion models ensure stability and fidelity in high-resolution terrain synthesis. Rather than replacing predictive models, generative approaches primarily enhance data quality, diversity, and uncertainty representation, thereby strengthening the robustness and generalization of landslide identification and forecasting frameworks.

3.4 Models for Anomaly Detection in Potential Landslide Identification

Anomaly detection provides a complementary perspective to supervised landslide classification by focusing not on what constitutes a landslide, but on when and where a slope begins to deviate from its normal state. In potential landslide identification, this paradigm is particularly valuable because catastrophic failures are often preceded by subtle, progressive, and spatially heterogeneous signals. Typical anomalies include unexpected acceleration in surface displacement, coherence loss in InSAR observations, or irregular fluctuations in multi-sensor monitoring data, which may emerge well before visible slope failure (Deijns et al., 2020; Jiang et al., 2020).

Compared with conventional anomaly detection approaches based on empirical thresholds or predefined statistical rules, deep learning-based methods offer a critical advantage: they can learn complex, nonlinear "normality patterns" directly from data, without requiring explicit assumptions about failure modes. This shift is especially important in landslide-prone environments, where background variability driven by rainfall, vegetation dynamics, and sensor noise often masks early instability signals. By modeling high-dimensional spatiotemporal dependencies, deep learning enables a more adaptive and context-aware identification of abnormal slope behavior.

AEs constitute the most widely adopted framework for unsupervised anomaly detection in landslide monitoring. Rather than explicitly detecting failures, AEs are trained to reconstruct normal system states, such as stable slope displacement time series or radar backscatter signatures (Sakurada and Yairi, 2014; Zhou and Paffenroth, 2017). When exposed to abnormal inputs (such as sudden deformation acceleration or coherence degradation) the reconstruction error increases, providing an implicit indicator of potential instability. This reconstruction-based logic is particularly attractive in landslide applications, where labeled failure data are scarce or incomplete. For instance, Shakeel et al. (2022) developed an InSAR deformation anomaly detector based on an AE–LSTM architecture. Experimental analyses using synthetic deformation test scenarios achieved an overall performance accuracy of 91.25%.

However, deterministic AEs implicitly assume that "normal" behavior can be represented by a single compact manifold, which may be insufficient for landslide systems characterized by multiple deformation regimes. VAEs address this limitation by explicitly modeling uncertainty in the latent space through probabilistic inference (Kumar et al., 2024; Pol et al., 2019). By learning a distribution rather than a single representation of normal slope behavior, VAEs are better suited to capture the intrinsic variability of environmental and geotechnical conditions (Kingma and Welling, 2013; Li et al., 2020; Park et al., 2018). Recent studies indicate that VAEs outperform conventional AEs when anomaly detection involves multivariate inputs combining displacement, rainfall, and hydrological factors, enabling a more robust identification of transitional instability stages (Nawaz et al., 2024; Han et al., 2025). Nevertheless, the probabilistic nature of VAEs also introduces practical challenges, including higher data requirements and the need for operationally meaningful thresholding strategies.

GANs offer an alternative perspective on anomaly detection by exploiting the discriminator's ability to differentiate between learned "normal" patterns and unfamiliar inputs (Kang et al., 2024; Xia et al., 2022). In landslide monitoring, GAN-based approaches learn the distribution of stable slope features, while deviations from this distribution are interpreted as anomalies (Radoi, 2022). Extensions such as AnoGAN further adapt this adversarial framework by explicitly embedding anomaly scoring mechanisms into the latent space (Lin et al., 2023; Thomine et al., 2023). While GAN-based methods have shown promise in detecting subtle deviations in complex data distributions, their training instability and sensitivity to hyperparameters remain practical limitations, particularly for operational early-warning systems.

Temporal models, including RNNs, LSTMs, and GRUs, play a distinct yet complementary role in anomaly detection by emphasizing when abnormal behavior emerges. These models learn expected temporal evolution patterns in displacement or rainfall time series and flag deviations from predicted trajectories (Zamanzadeh Darban et al., 2024; Zhang et al., 2022a). In landslide early-warning scenarios, this temporal sensitivity is crucial for identifying acceleration phases rather than static anomalies. Hybrid architectures that integrate temporal models with AEs or GANs further enhance anomaly detection by jointly capturing spatial reconstruction errors and temporal inconsistencies, enabling multi-source consistency checks across monitoring networks. For instance, Geiger et al. (2020) demonstrated a growing trend of utilizing LSTM networks as both the generator and discriminator within GAN frameworks for time-series anomaly detection. Similarly, Whitaker (2023) illustrated the application of LSTM–GAN architectures in identifying temporal anomalies.

Deep learning-based anomaly detection shifts landslide identification from static classification toward dynamic state monitoring, making it particularly suitable for early recognition of slope instability under evolving environmental conditions. Although these methods do not directly predict landslide occurrence, they provide an essential early-warning layer by highlighting abnormal system behavior that warrants further physical interpretation or intervention.

3.5 Models for Data Fusion in Potential Landslide Identification

In practical applications, the identification of potential landslide hazards is a complex task that influences by multiple factors (Zhang et al., 2018). These factors are often reflected through different data sources. We can roughly divide heterogeneous data into four categories: image data, time series data, structured data, and textual data. Given this heterogeneity, data fusion is essential for the accurate identification of potential landslides (see Fig. 5).

Conventional data fusion approaches in landslide studies (such as feature concatenation, weighted linear combination, or statistical multivariate analysis) generally rely on predefined assumptions regarding variable independence or linear interactions. While these methods are computationally efficient, they struggle to capture the nonlinear, scale-dependent, and cross-modal relationships that characterize real-world landslide processes. In contrast, deep learning–based data fusion models provide a data-driven means to automatically learn high-order feature interactions across heterogeneous inputs, thereby offering a more flexible and expressive framework for potential landslide identification.

Among existing architectures, Graph Neural Networks (GNNs) have attracted increasing attention due to their ability to explicitly represent non-Euclidean spatial relationships. Landslide-related terrain units (e.g. slope units, grid cells, or monitoring stations) are inherently interconnected through topography, hydrological pathways, and geological continuity (see Fig. 5b). Conventional CNN-based fusion models, which operate on regular grids, are limited in capturing such irregular spatial dependencies. By contrast, GNNs represent spatial entities as nodes and their geospatial, hydrological, or geological relationships as edges, enabling the propagation of information across topologically connected units (Scarselli et al., 2008; Ying et al., 2018; Zeng et al., 2022).

In landslide identification and forecasting, this graph-based representation allows geomorphic and hydrological signals to be explicitly transmitted between adjacent or functionally connected units, thereby better reflecting slope interaction mechanisms. For example, Kuang et al. (2022) proposed an innovative landslide forecasting model based on GNNs, in which graph convolutions are employed to aggregate spatial correlations among different monitoring sites. Ren et al. (2025) introduced a novel GNN framework with conformal prediction (GNN-CF) for landslide deformation interval forecasting, addressing the limitations of conventional models in handling predictive uncertainty.

According to the differences in message passing and aggregation methods, GNNs have derived various variants. For example, Graph Convolutional Network (GCN) is generated by generalizing the convolutional operation to graph-structured data (Kipf and Welling, 2016; Sharma et al., 2022; Wang et al., 2020a), and Graph Attention Network (GAT) dynamically weights the importance of neighboring nodes by introducing the attention mechanism (Veliˇckovi´c et al., 2017; Yuan et al., 2022; Zhou and Li, 2021). The emergence of these new architectures makes GNN variants more targeted than conventional GNNs and suitable for modeling heterogeneous relationships. Currently, they are often used for weighted analysis of the impacts of different geographical factors on landslides (Kuang et al., 2022; Li et al., 2025; Zhang et al., 2024e).

Beyond graph-based models, Transformer architectures have emerged as a unifying framework for multimodal data fusion in landslide studies. As highlighted in Section 3.2, the Transformer's self-attention mechanism and modular architecture make it a universal framework for processing sequential data and enabling multimodal fusion (see Fig. 5c).

In this context, the core advantage of the Transformer lies in its ability to integrate diverse input data (e.g., satellite imagery, GPS time series, and geological maps). It achieves this by employing independent embedding layers to convert each modality into a unified vector representation, which is then fused through the self-attention mechanism. This mechanism computes the interactions and correlations among all elements across different modalities, thereby enabling the model to capture cross-modal dependencies and extract joint feature representations within a unified framework. This capability makes the Transformer particularly suitable for landslide studies (Li et al., 2025). For example, Piran et al. (2024) enhanced short-term precipitation forecasting by applying transfer learning with a pre-trained Transformer model. Zhang et al. (2024e) incorporated Transformer modules to build a graph-Transformer model that integrates global contextual information for the generation and analysis of Landslide Susceptibility Maps (LSMs).

In conclusion, deep learning–based data fusion provides a flexible and unified framework for integrating heterogeneous landslide-related data, including spatial, temporal, and topological information. By enabling joint representation learning across multiple data modalities, fusion-oriented architectures such as GNNs and Transformers have substantially enhanced the capability of potential landslide identification to capture complex environmental interactions that cannot be adequately represented by single-source or loosely coupled models. As a result, data fusion has become a critical methodological component in contemporary deep learning–based landslide hazard studies.

Comment #5:

5. The captions of Figures 2-5 should clearly indicate their function rather than repeat text descriptions.

Response:

- Thank you for this helpful suggestion. We agree that the captions should concisely state the figure's purpose. Accordingly, we have revised the captions for Figures 2-5 to clearly articulate their demonstrative or illustrative function within the manuscript. The updated captions now highlight what each figure is intended to show, rather than describing its content in a manner that overlaps with the main text.

Original Description in Figure 2

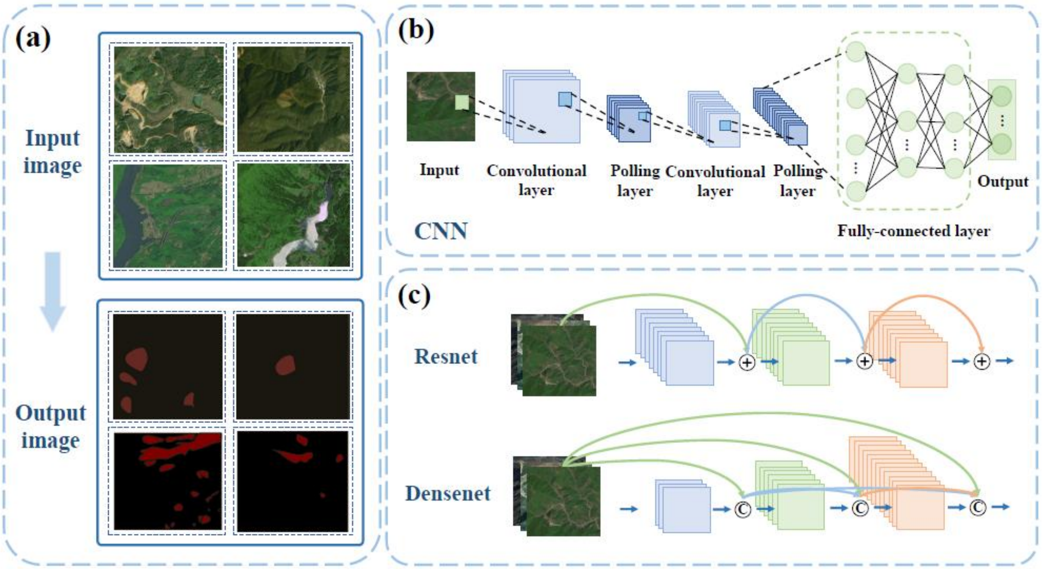

Figure 2. The role of deep learning models in image analysis and processing. (a) Comparison of landslide images before and after identification. (b) Schematic of a basic CNN architecture. A conventional CNN typically comprises stacked convolutional layers, pooling layers, and fully connected layers. (c) Comparative schematic of ResNet and DenseNet architectures. In contrast to ResNet, which combines features through summation before passing them to subsequent layers, DenseNet integrates features via channelwise concatenation.

Revised Description in Figure 2

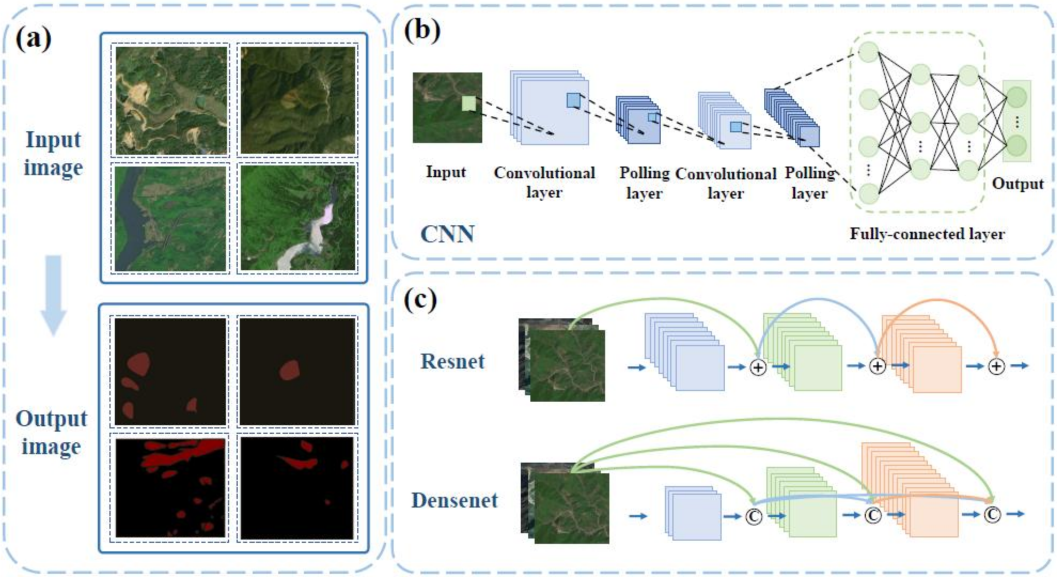

Figure 2. Functional pipeline of CNN-based models for image analysis and processing. (a) Semantic mapping process: demonstrating the transition from optical input to binary classification for target identification. (b) Segmentation performance: visualizing the model’s capability to delineate precise landslide boundaries (binary masks) from optical imagery. (c) Optimization strategies: comparing skip-connections and dense connectivity for enhancing gradient flow and feature reuse.

Original Description in Figure 3

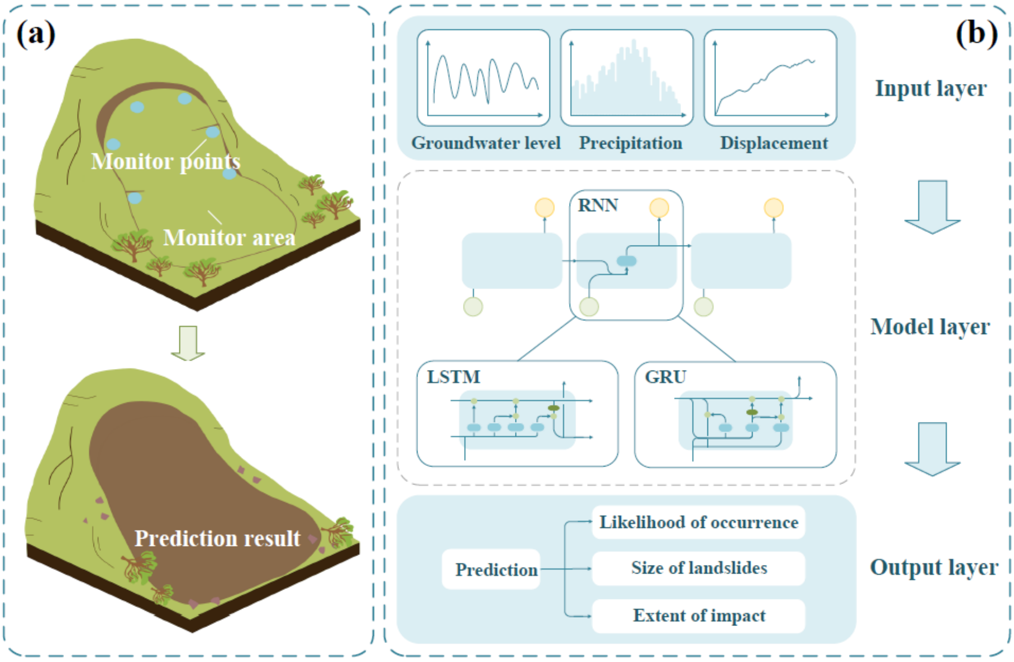

Figure 3. The role of deep learning models in time series analysis. (a) In potential landslide identification, time series data can be obtained through monitoring. (b) RNNs, LSTMs, and GRUs provide more accurate information for landslide prediction by processing time series data.

Revised Description in Figure 3

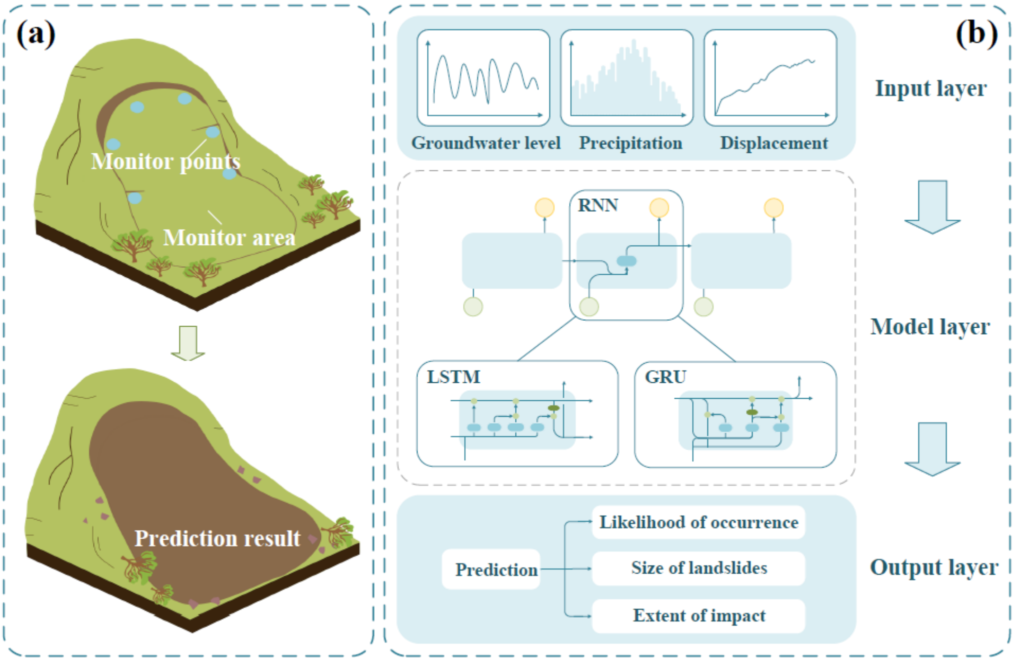

Figure 3. Analytical framework of RNN-based models for time series analysis. (a) From field monitoring to predictive insight: outlining the transformation of multi-source field monitoring data into predictive landslide intelligence. (b) Processing temporal dependencies: illustrating the recursive logic of RNN, LSTM, and GRU in processing sequential variables.

Original Description in Figure 4

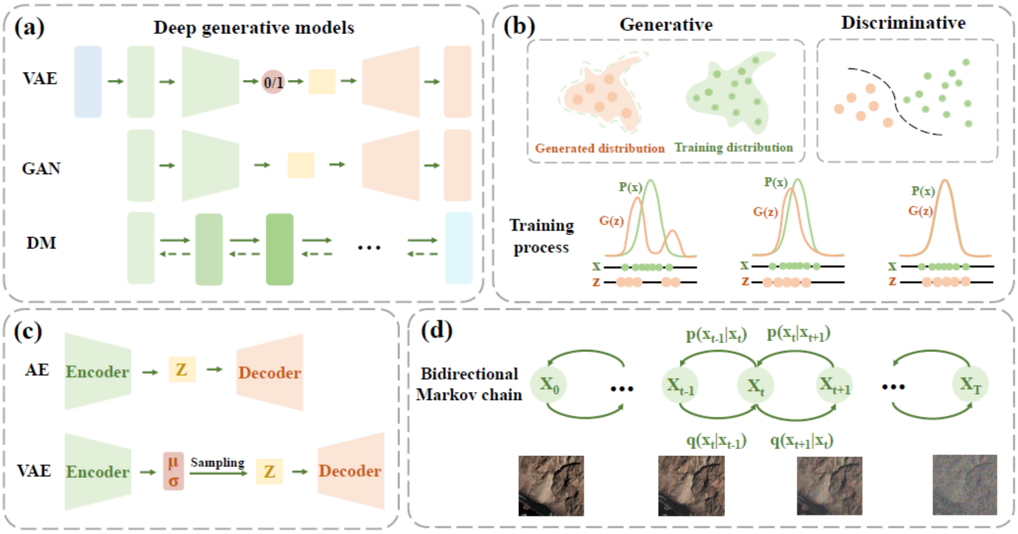

Figure 4. The role of deep learning models in data generation. (a) Comparative schematic of three commonly used deep generative model architectures. GAN: adversarial training. VAE: maximize variational lower bound. Diffusion models: gradually add Gaussian noise and then reverse. (b) Schematic of the adversarial training workflow for GAN-based data generation. (c) Comparative architecture of AE and its variational counterpart, VAE. (d) Schematic of a diffusion model applied to denoise potential landslide data.

Revised Description in Figure 4

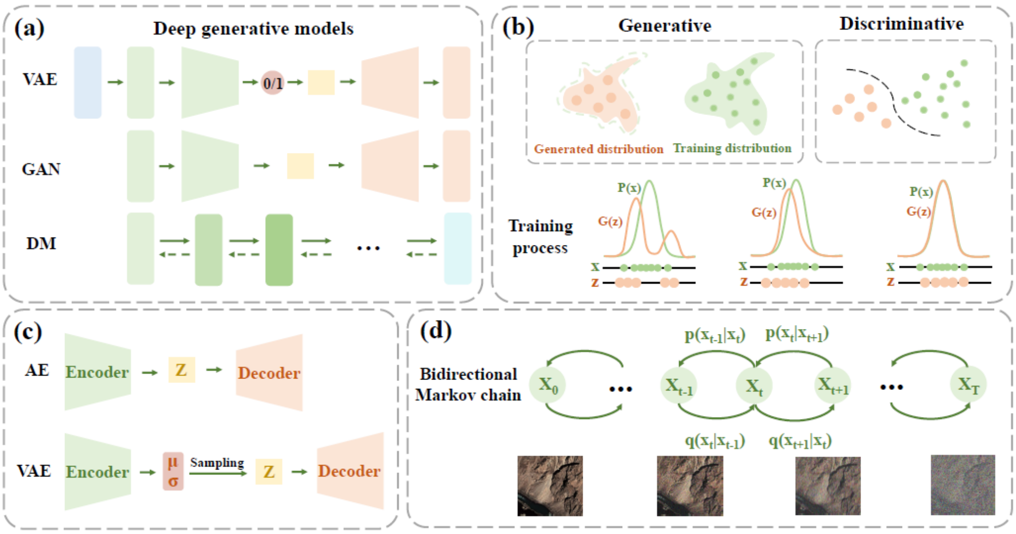

Figure 4. Comparative mechanisms of deep generative models for data generation. (a) Contrasting fundamental training objectives: VAE (maximizing variational lower bounds), GAN (adversarial gaming), and Diffusion models (iterative noise reversal). (b) Adversarial learning: function of the generator-discriminator competition in improving sample fidelity. (c) Latent space modeling: highlighting the probabilistic sampling layer in VAEs that enables diverse sample generation compared to standard AEs. (d) Iterative denoising: the mechanism of reconstructing high-resolution imagery through reverse diffusion.

Original Description in Figure 5

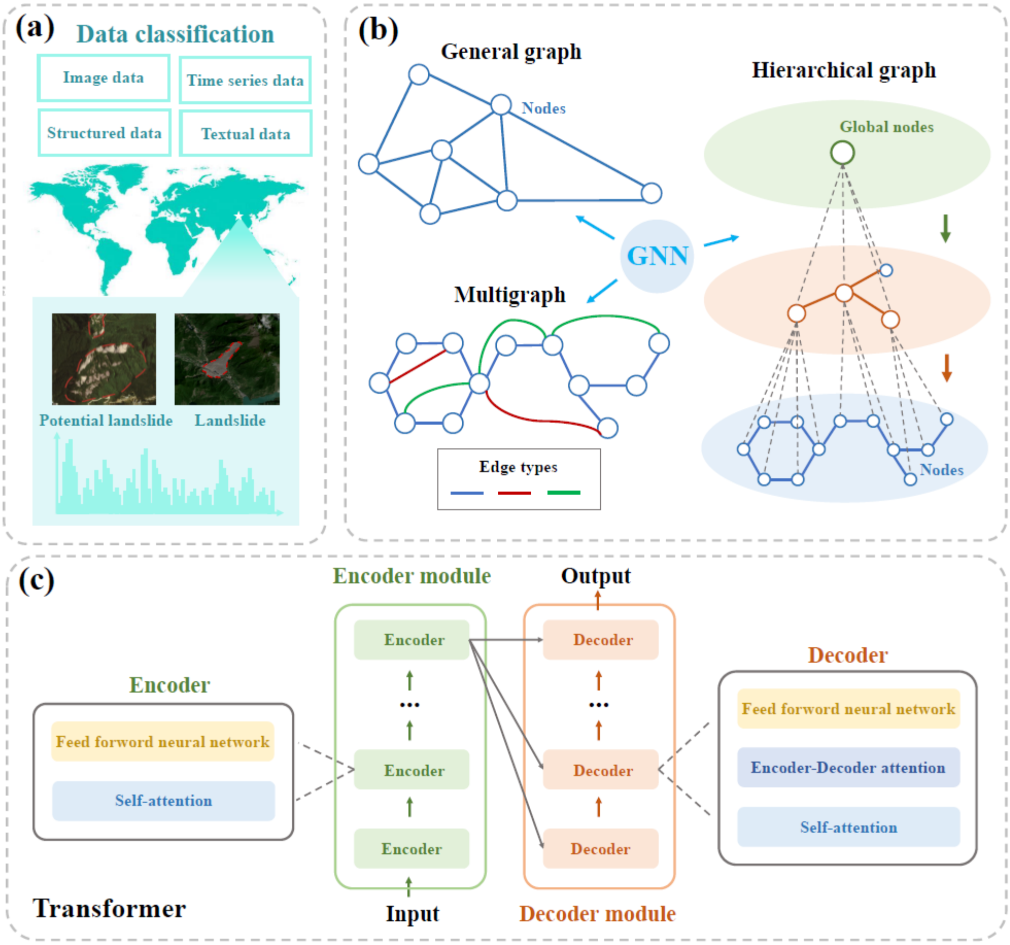

Figure 5. The role of deep learning models in data fusion. (a) Classification of heterogeneous data for potential landslide identification. (b) Schematic of general graph and more complex graphs. (c) Schematic of the fundamental Transformer architecture.

Revised Description in Figure 5

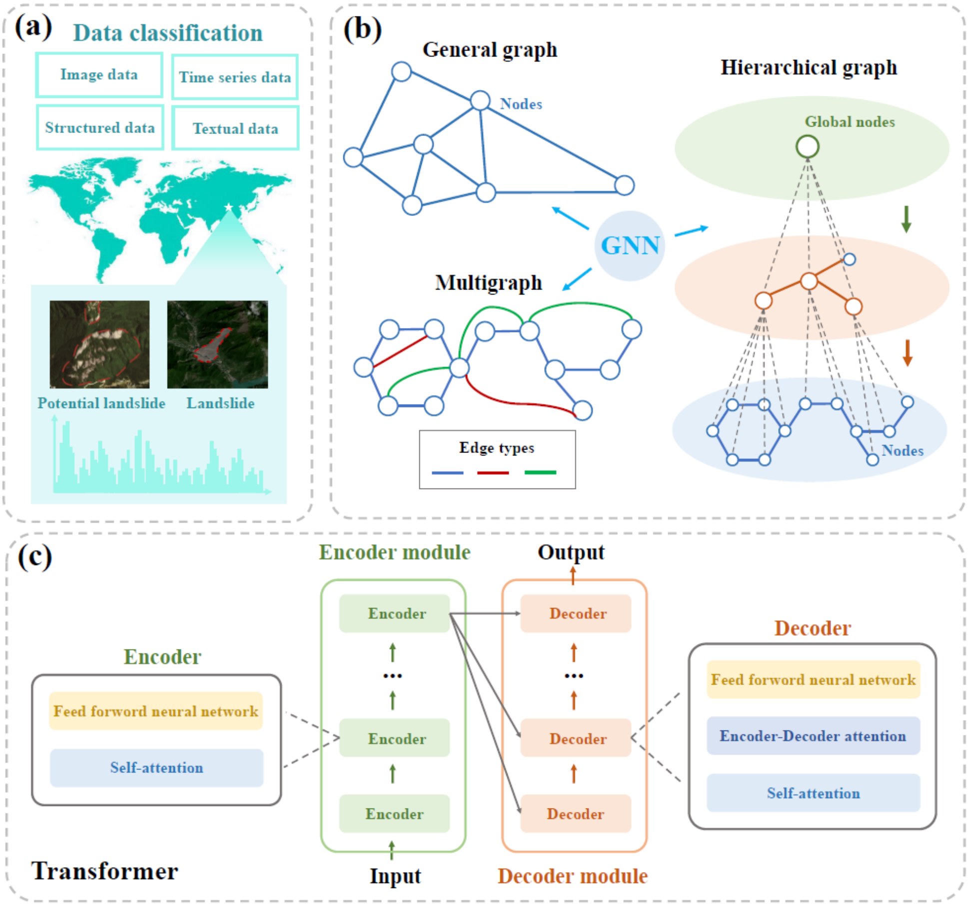

Figure 5. Integrated framework of GNNs and Transformers for data fusion. (a) Multi-source integration: the architectural flow for synthesizing heterogeneous datasets (spatial images, time-series, and structured data) to support robust decision-making. (b) Topology modeling: GNN mechanisms designed to aggregate spatial dependencies across general, multi-graph, and hierarchical slope networks. (c) Global contextual attention: the Transformer architecture utilizing self-attention mechanisms to capture long-range dependencies in sequence-based or flattened spatial features.

Comment #6:

6. Several minor word-joining problems occur due to PDF line merges. Please check the entire manuscript carefully and correct spacing between words.

Response:

- Thank you for your careful reading and for highlighting this formatting issue. We have carefully re-checked the entire manuscript and corrected all instances of unintended word-joining or irregular spacing introduced during line breaks. Please see the revised version below.
- These corrections have been implemented in the revised source file and are consistently reflected in the updated PDF version.
- Once again, we thank you for your careful review and valuable feedback!

Original Description in the Manuscript

……

With the rapid advancement of UAVs, centimeter-level vertical and oblique aerial photogrammetry is now achievable (Fan et al., 2020). The high-definition cameras mounted on UAVs are able to capture the subtle cracks on the surface of the mountain.These cracks may be early signs of a landslide (Sun et al., 2024a). By conducting a comparative analysis of the images taken at different times, the development and changes of the cracks can be monitored, including the increase in the length, width and depth of the cracks, as well as the changes in the crack orientation.

……

For example,Long et al. (2018) proposed a GBSAR persistent scatterer point selection method based on the mean coherence coefficient, amplitude dispersion index, estimated signal-to-noise ratio, and displacement accuracy index.

…….

Representative examples include the CAS Landslide Dataset, a large-scale, multi-sensor dataset explicitly designed for deep learning-based landslide mapping (Xu et al., 2024); the Landslide4Sense (L4S) benchmark, developed within an international competition, which provides multisource satellite image patches (Ghorbanzadeh et al., 2022b); and the Diverse Mountainous Landslide Dataset (DMLD), which emphasizes high-resolution instances from complex mountainous terrains (Chen et al., 2024b).

…….

For example, Feng et al. (2024) achieved the first implementation of using a GAN to generate synthetic high-quality landslide images, aiming to address the data scarcity issue that undermines the performance of landslide segmentation models.

……

To provide a systematic and targeted analysis, thissection organizes the applications according to four major triggering categories: rainfall-induced landslides, earthquake-induced landslides, human activity-induced landslides, and multi-factor-induced landslides (see Fig. 6).

……

Direct fusion of such multi-modal data induces feature space incompatibility, hindering cross-modal correlation extraction (Cai et al., 2021;Jin et al., 2022).

……

This dynamic progression across different timescales creates a fundamental modeling challenge: since the numerical simulation of long-term creep requires a long time step, while the dynamic process of short-term abrupt changes requires a time resolution in the microsecond level, it is difficult to establish a unified model for these two situations. This will further intensify the conflict of time scales.

……

Furthermore, the spatial analysis capabilities of Geographic Information System (GIS) were integrated into the practical identification workflow, enabling the study area to be partitioned into distinct landslide risk categories.This risk stratification considers the combined influence of region-specific factors, ensuring scientifically robust and practically viable classifications.

……

Revised Description in the Manuscript

……

With the rapid advancement of UAVs, centimeter-level vertical and oblique aerial photogrammetry is now achievable (Fan et al., 2020). The high-definition cameras mounted on UAVs are able to capture the subtle cracks on the surface of the mountain. These cracks may be early signs of a landslide (Sun et al., 2024a). By conducting a comparative analysis of the images taken at different times, the development and changes of the cracks can be monitored, including the increase in the length, width and depth of the cracks, as well as the changes in the crack orientation.

……

For example, Long et al. (2018) proposed a GBSAR persistent scatterer point selection method based on the mean coherence coefficient, amplitude dispersion index, estimated signal-to-noise ratio, and displacement accuracy index.

……

Representative examples include the CAS Landslide Dataset, a large-scale, multi-sensor dataset explicitly designed for deep learning-based landslide mapping (Xu et al., 2024); the Landslide4Sense (L4S) benchmark, developed within an international competition, which provides multi-source satellite image patches (Ghorbanzadeh et al., 2022b); and the Diverse Mountainous Landslide Dataset (DMLD), which emphasizes high-resolution instances from complex mountainous terrains (Chen et al., 2024b).

…….

For example, Feng et al. (2024) achieved the first implementation of using a GAN to generate synthetic high-quality landslide images, aiming to address the data scarcity issue that undermines the performance of landslide segmentation models.

……

To provide a systematic and targeted analysis, this section organizes the applications according to four major triggering categories: rainfall-induced landslides, earthquake-induced landslides, human activity-induced landslides, and multi-factor-induced landslides (see Fig. 6).

……

Direct fusion of such multi-modal data induces feature space incompatibility, hindering cross-modal correlation extraction (Cai et al., 2021; Jin et al., 2022).

……

This dynamic progression across different timescales creates a fundamental modeling challenge: since the numerical simulation of long-term creep requires a long-time step, while the dynamic process of short-term abrupt changes requires a time resolution in the microsecond level, it is difficult to establish a unified model for these two situations. This will further intensify the conflict of time scales.

……

Furthermore, the spatial analysis capabilities of Geographic Information System (GIS) were integrated into the practical identification workflow, enabling the study area to be partitioned into distinct landslide risk categories. This risk stratification considers the combined influence of region-specific factors, ensuring scientifically robust and practically viable classifications.

……

- With the completed clarifications and revisions, we hope to have effectively responded to all the issues you pointed out. Thank you deeply for your thoughtful consideration and guidance!
- Have a nice day!
- Pan Jiang & Zhengjing Ma & Gang Mei